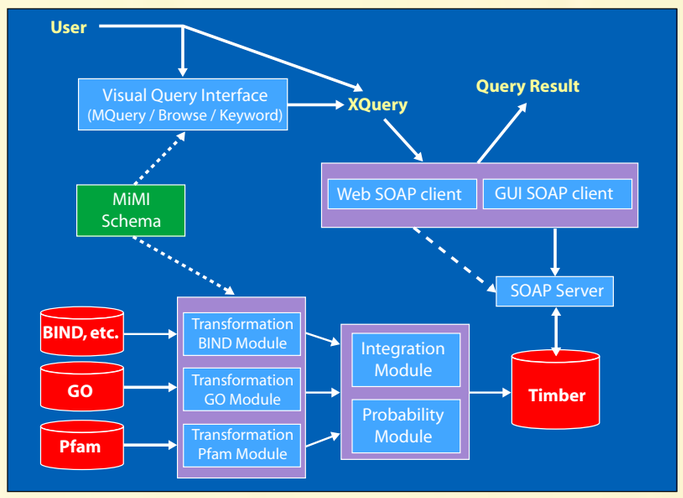
MiMI: Michigan Molecular Interactions

Adriane Chapman, Magesh Jayapandian, Cong Yu, Glenn Tarcea, H. V. Jagadish
University of Michigan, Ann Arbor, Michigan, USA

Overview

- Goal: Effective integration and retrieval of biological data from multiple sources
- Deep Merge
  - Integrates heterogeneous and complimentary data sources
  - Facilitates provenance-tracking and supports probabilistic manipulation of unreliable data
- Query Interface
  - Supports standard query interfaces: Keyword, Browse, and XQuery
  - MQuery: A flexible query interface providing functionalities with minimum user effort
  - NaLIX: A Natural Language Interface to Query XML
- Features
  - Provenance Tracking
  - Probability Manipulation
  - Homologue-based Interaction Prediction
- Available At: http://mimi.ctaalliance.org

System Architecture

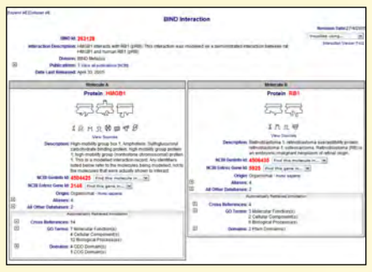


Acknowledgements

- University of Michigan
  - Medical School: Philip Andrews; Brian Athey; David States
  - Department of Electrical Engineering and Computer Science: Jignesh Patel
  - Department of MCDB: Anuj Kumar
  - School of Information: Dragomir Radev
  - Department of Linguistics: Steve Abney
- Wayne State University
  - Russ Finley
- Thanks to the following groups for sharing their data
  - Samuel Lunenfeld
  - Research Institute,
  - Cold Spring Harbor Laboratory,
  - European Bioinformatics Institute
  - Center for Cancer Systems Biology at Harvard
  - Max Delbueck Center
  - EMBL
  - Gene Ontology Consortium
  - Johns Hopkins University / Institute of Bioinformatics
  - Mt. Sinai Hospital / Blueprint
  - NCBI
  - Sanger Institute
  - UCLA
- Support
  - NSF grant IIS-0219513, NIH project 1-U54-DA021519-01A1, NSF through a graduate fellowship

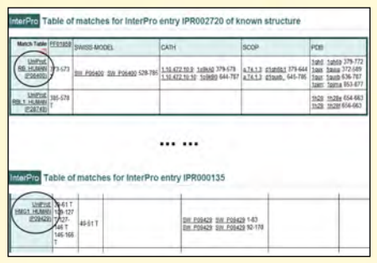
Deep Merge

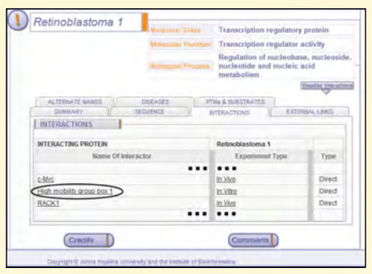
Heterogeneous Sources



BIND

InterPro





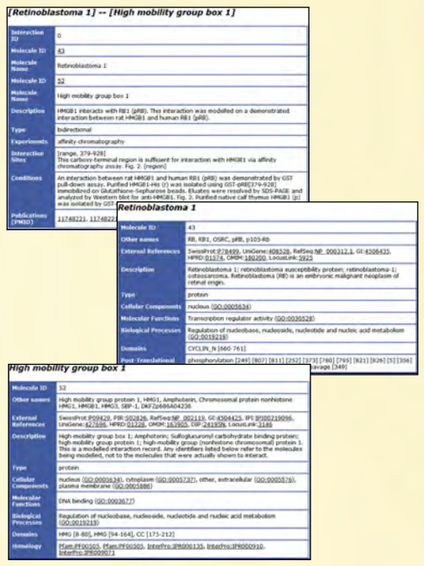
HPRD

Merging Guidelines

- Preserve all data
- Maintain provenance
- Utilize an identity function to determine whether two objects are fundamentally the same in the real world and should be consolidated

- 9,860 Center for Cancer Systems Biology Records
- 287,072 BIND Records
- 94,533 HPRD Records
- 73,561 DIP Records
- 69,065 GRID Records
- 130,622 IntAct Records
- 5,178 Max Delbrueck Center Records
- 89,662 Reactome Records

MiMI Records



Statistics

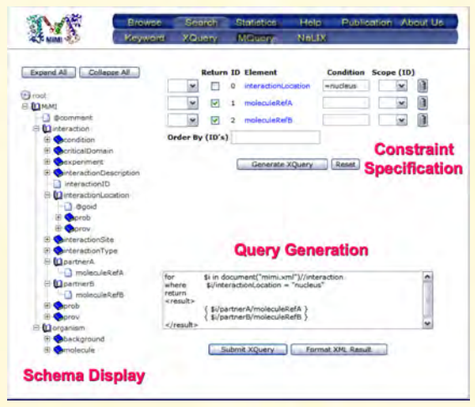
- Sources
  - Interaction Data: Center for Cancer Systems Biology Dataset, BIND, DIP, HPRD, GRID, IntAct, Max Delbrueck Center Dataset
  - Pathyway Data: Reactome
  - Auxiliary Data: GO, OrganelleDB, OrthoMCL, Pfam, ProtoNet, miBlast, InterPro, IPI

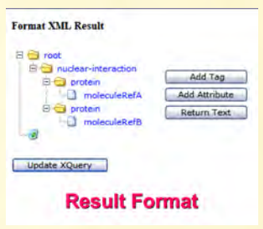
Query Interface

Components

- Traditional
  - **Browse:** Enables users to trawl through records of interest
  - **Keyword:** Facilitates easy lookup of a particular record
  - **XQuery:** Permits users to write their own sophisticated XQuery
- MQuery
  - Allows users to point-and-click queries
  - Teaches users by example how to use XQuery
- NaLIX
  - An extension that allows users to write natural language queries against the database

MQuery Interface





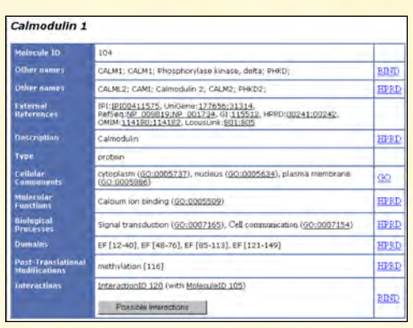
- Any web browser can be used to load the interface
- Supported query operations include selection, projection, join, and sorting

MQuery Advantages

- Simplicity
  - Users pose declarative queries in XQuery without prior knowledge of
    - XQuery syntax
    - Schema of the database
- Expressivity
  - MQuery is more expressive than traditional (static) form-based interfaces which are typically ad hoc and non-extensible
- Compatibility
  - MQuery is HTML-based and can be loaded and viewed on any standard web-browser

Features

Provenance and Probability

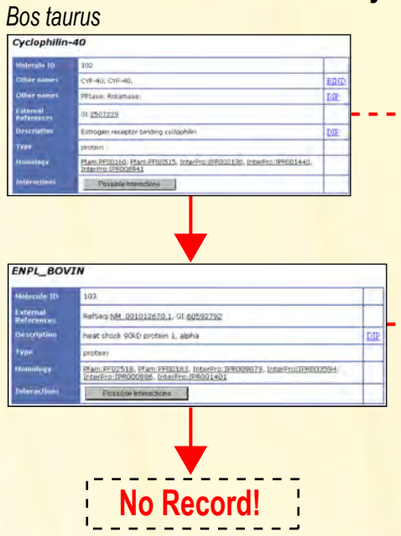


- Provenance
  - Tracks data origin
  - Describes manipulations being performed on the data
  - Enables users to decide which sources they trust and wish to retrieve
- Probability
  - Utilizes reported false positive/negative rates of experiments as base probabilities
  - Deduces the probability of an interaction based on base probabilities using standard probability model
  - Allows users to adjust the default probabilities of given experiments or interaction

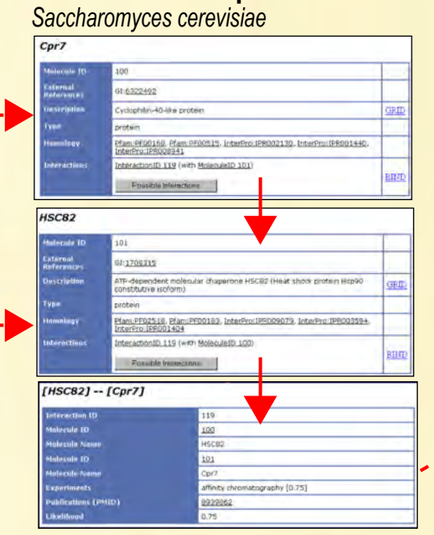
Interaction Prediction

Q: Does Cyclophilin-40 interact with Hsp90 in *Bos taurus* ?

Bos taurus



Saccharomyces cerevisiae



Interaction exists in *S. cerevisiae*
A literature search discovered a reported interaction between homologous proteins Cyclophilin-40 – Hsp90 in *B. taurus*.